AN E-MAIL SERVER-BASED SPAM FILTERING APPROACH

MUMTAZ MOHAMMED ALI AL-MUKHTAR
College of Information Engineering, AL-Nahrain University
IRAQ

ABSTRACT

The spam has now become a significant security issue and a massive drain on financial resources. In this paper, a spam filter is introduced, which works at the server side. The proposed filter is a combination of antispam techniques. The integrated solution create a spam filtering system which is more robust and effective than each of the comprising techniques. The task of proposed filter is to minimize the ability of the spammers to distract the network by the spam. That is done by blocking the spam messages at the server level. A server-based solution is normally more advantageous than protecting e-mail users individually. Such a solution gives more control to administrators.

Keywords*: e-mail spam, spam filtering, machine learning, blacklisting, RBL.*

1. INTRODUCTION

In the past few years, Internet technology has affected our daily communication style in a radical way: the electronic mail (e-mail) concept is used very extensively for communications nowadays. This technology makes it possible to communicate with many people simultaneously in a way so easy and cheap that it is currently considered the first worldwide medium into business sector [1].

However, the abuse of e-mails has the drawback that the volume of e-mails that show up in mailboxes has exponentially increasing. Moreover, many e-mails are received by users without their desire: “spam mail” (or “junk mail” or “bulk mail”) is the general name used to denote these types of e-mail. Spam mails, by definition, are the electronic messages posted blindly to thousands of recipients, usually for advertisement, and represent one of the most serious and urgent information overload problems [2].

Spam has caused some serious problems. Firstly, it wastes a mass of network resources that are very important for network users, especially those in enterprises or corporations. People need to spend a lot of time to deal with spam every day. Even worse, many current spam mails bring users unexpected malicious attachments which would seriously crack the user’s system. Therefore, spam is a headachy problem [3].

Spam filtering (i.e., distinguishing between spam and legitimate e-mail messages) is a commonly accepted technique for dealing with spam [4]. Spam filters vary in functionality from black-lists of frequent spammers to content-based filters. The latter are generally more powerful, as spammers often use fake addresses. Existing content-based filters search for particular keyword patterns in the messages. These patterns need to be crafted by hand, and to achieve better results they need to be tuned to each user and to constantly maintained [5].

These content-based approaches include statistical classification [6,7,8,9], rule–based filtering [10] and neural networks-based solutions [11]. Other classes of filtering include challenge-response [12], ontology driven filter [13,], collaborative approach [14], and using visual features for spam filtering [15].

The objective of this paper is working on the server side. All users’ mail is filtered by a central server, and that server keeps track of user’s profile. The cost of purchasing software to protect users individually can be higher than protecting them indirectly by protecting the server. Server-based solutions give administrators more control. Even if a company purchases anti-spam software for all of its employees, it is not guaranteed that they will use it correctly. Furthermore, employees who do want to benefit from their anti-spam software will have to spend time tuning their spam filters. Some might not tune them correctly; therefore, spam messages will continue to appear in their mailbox or, ever worse, legitimate e-mails could be lost.

As spammers become more sophisticated, the tokens found in the message body used by text-based filters to distinguish between spam and legitimate messages will no longer be sufficient. Thus, the attention is to include other characteristics that spammers are unable to successfully obscure. A multilayer consistent solution has been proposed to overcome the spam problem and spammers effort to hide their track. The proposed spam detection system decides whether the received e-mail is spam or not by using these information: IP address of the sending server, real-time black-hole list, domain name of the sender, behavior of the sender, and the body of the e-mail transferred. The advantages of these techniques have been taken to help enhance spam filtering.

2. A PROPOSED FILTERING SPAM STRUCTURE

The proposed filter concerns with an e-mail server side. The e-mail servers have different work features than other e-mail parts. They can connect with other servers to receive the incoming messages and get the resource IP of the delivering servers, thus the filter can check the source whether it has been trusted or not.

The proposed filter consists of many stages as shown in figure 1. Each stage has its special mechanism to handle the spam. The following sections will describe the function of each stage.

IP ADDRESS BLACKLIST

Blocking the e-mail from certain domains known to be used by spammers can yield good results. It is a simple mechanism to stop the spam by the sender IP address. Every connection that has an unaccepted IP address will be considered as a spam mail.

The receiving server will get the IP address of the sending mail server from the SMTP HELO command, and checks it against the IP addresses in the Blacklist. If a match is found the sender will be considered a spammer and the connection will be disabled, otherwise the filter passes the mail to the next mechanism.

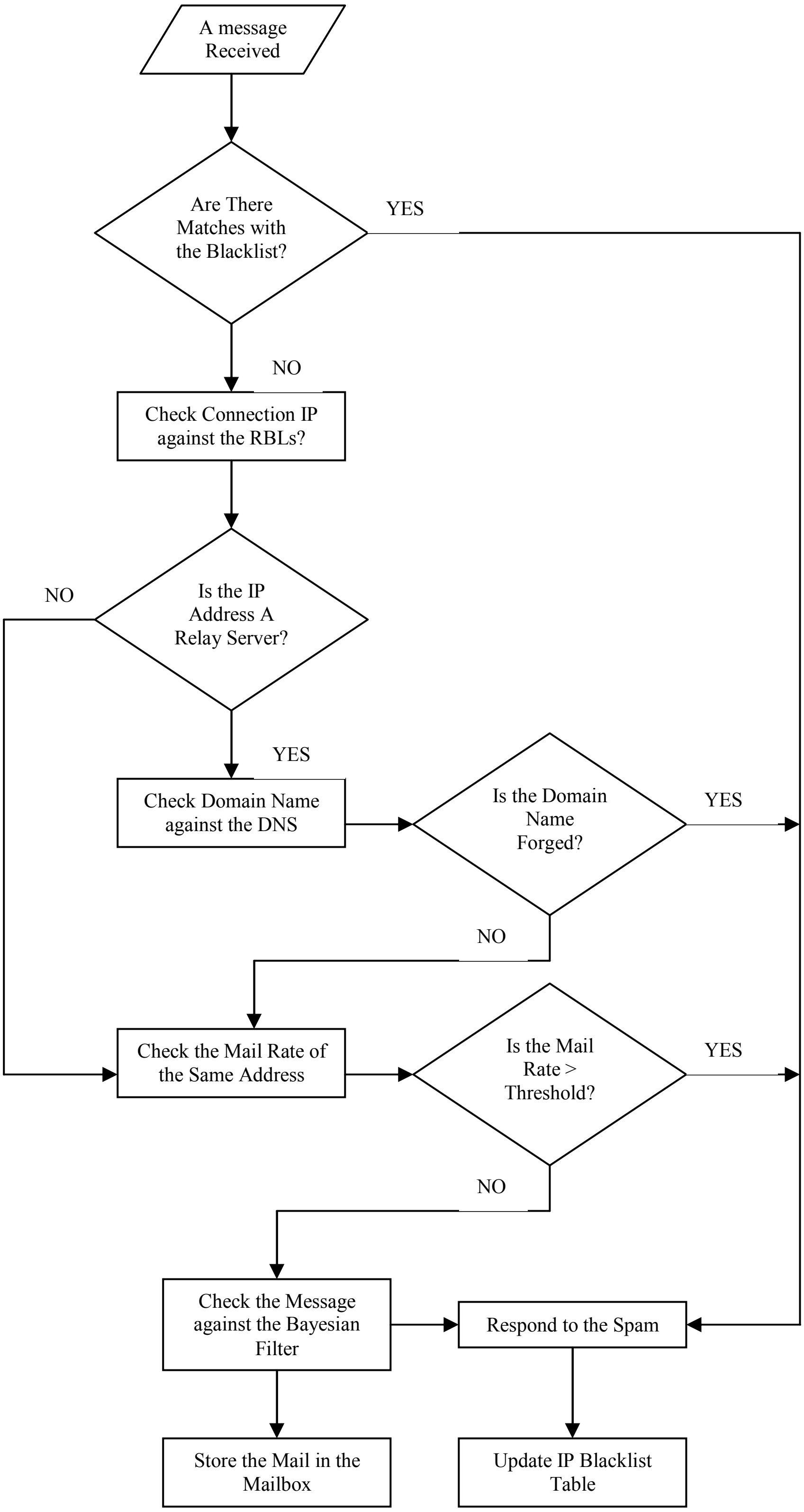
A message Received
Are There Matches with the Blacklist?
YES
NO
Check Connection IP against the RBLs?
Is the IP Address A Relay Server?
NO
YES
Check Domain Name against the DNS
Is the Domain Name Forged?
YES
NO
Check the Mail Rate of the Same Address
Is the Mail Rate > Threshold?
YES
NO
Check the Message against the Bayesian Filter
Respond to the Spam
Store the Mail in the Mailbox
Update IP Blacklist Table

Figure (1) The Proposed Filter Flow Control

2.2 REAL TIME BLACKHOLE LIST (RBL)

This technique, commonly referred to as RBL (real-time black-hole lists), checks the incoming IP address against Black Lists to verify that the sending server is not listed as an open mail relay that spammers can use to relay their unsolicited e-mails. The RBL contains a list of open relay IP addresses maintained by third-party organizations. One of the most reliable databases of server addresses, is maintained by ORDB.org [16].

The IP address of the sender is extracted and checked with the RBL. If the sender IP address is an open relay server, then the filter will consider the sender as a not trusted one. Here the filter will send the IP address to the DNS Lookup to check whether it has been forged or not. If there has been no positive result from the RBL, that is the IP address has not been an open relay server, the filter considers the sender as trusted and its address is real. Nevertheless the filter will send the IP address to the mail rate control mechanism.

2.3 DNS LOCKUP

This technique verifies that the domain name of the sender has not been spoofed. The proposed filter extracts the domain name from the "from:" field of the address or sender ID-address. The receiving server will get the host name of the sending mail server from the "from:" filed of the header or sender ID-address, performs a simple domain name server (DNS) query and compares the connected IP address with the retrieved IP addresses list to check if there is a match with an IP of the retrieved IP address list.

If the domain name had been forged, then the proposed filter will consider the message or the connection as a spam, and will reject this connection. Thereafter it adds this IP address to the IP blacklist. Else if the domain name has been correct, the filter will implement the mail rate control mechanism as the next stage.

This technique can identify whether the sending mail server is a legitimate one and has a valid host name. This will eliminate the majority of spam sent by mail servers connected to the Internet using a dial-up connection, as well as most ADSL and cable connections, simply because they are not registered in any domain name server as a qualified host.

2.4 MAIL RATE CONTROL

The proposed filter checks the behavior of the sender to stop who is trying to send a huge number of mails.

Rate mail controls can allow only a certain number of connections from the same e-mail address during a specified time. For example, a rate control time can be set of to 30 minutes with only a certain number of connections to be allowed in that given time period. If the administrator sets this parameter to 50 connections, this stage will block any correspondence after the first 50 connections that come from a single e-mail address within a given 30 minute time period.

The proposed filter also considers the "to:" field as input through the rate mail stage, because the sender can put many recipients' addresses in this field in one message.

2.5 BAYESIAN FILTER

The proposed filter uses probabilistic reasoning to decide whether or not a message is spam. This filter bases its choices on the Baye's rule, which is useful for calculating the probability of one event when one knows another event is true. In our case, the rule is used to determine the probability that an e-mail is spam given that it contains certain words. What makes Bayesian filters different from other filters is that they learn. To decide the probability that an e-mail is spam based on the words that it contains the filter needs to know about the e-mails that a user receives.

For the implementation of the Bayesian filter it is required to learn with a set of labeled messages. There are two stages carried out by the Bayesian filter: Training Level and Testing Level.

2.5.1 TRAINING LEVEL

This level is called training or learning level. This level is focused on gathering the information, concerning both spam and legitimate mails. At this stage the filter extract the tokens (words) of the labeled mail by an operation called tokenization that is responsible on extracting tokens from the mails, and storing them in tables. Two tables will be used, one for tokens of spam mails and other for tokens of legitimate mails. When an e-mail is declared as a spam, the spam table is updated by incrementing the frequency counts for each word contained in that e-mail. Legitimate mail counts are incremented similarly. The count number of spam and non spam e-mails is also recorded for use in the test level. We can get a list of spam mails from some dependable location in the web to learn filter with it. Also the unlabeled message when it is labeled by the filter will be considered as input to learn with at the test level.

2.5.2 TESTING LEVEL

In the test level the collected information about spam and non spam will be used as vectors to find the probability that the incoming mail is spam or not.

This process is implemented by the following steps:

i. Split e-mail in tokens.
- Need number of messages for spam and legitimate.
- Need frequency of each word for each type.
- Calculate probabilities
  - ➢ P(legitimate) = word frequency/number of legitimate messages.
  - ➢ P(spam) = word frequency/number of spam messages.
  - ➢ Calculate likelihood of being spam (spamicity) using a special form of Bayes' Rule where likelihood = a/(a+b), where a is

the probability of a legitimate word and b is the probability of spam word.

- Choose tokens whose combine probability is farthest from 0.5 either way. This is because the farher it is from 0.5 (neutral), with more certainty we can say it belongs to either strategy.

ii. Do this for n numbers for instance choose to have 15 extremes.

iii. Combine their probability to get a figure for message using Bayes' Rule

$$\frac{a\ b\ c}{a\ b\ c + (1 - a) * (1 - b) * (1 - c)}$$

If the end result is closer to 1.0, then the message is classified as spam, and if it is closer to 0.0, the message is classified as legitimate. The cutoff range we have specified for spam is that it should be greater than 0.85, but experimental results showed that most spam results are above 0.98.

3. FILTER EFFICIENCY AND RESULTS

The detection of spam introduces two sources of misclassification: false positive where a non spam e-mail is classified as spam and false negatives where spam slips through incorrectly identified as non spam.

Filter efficiency depends fundamentally on two factors. The first factor is spam detection percentage, and the second factor is the misclassification percentage. The proposed filter efficiency is observed against these two factors.

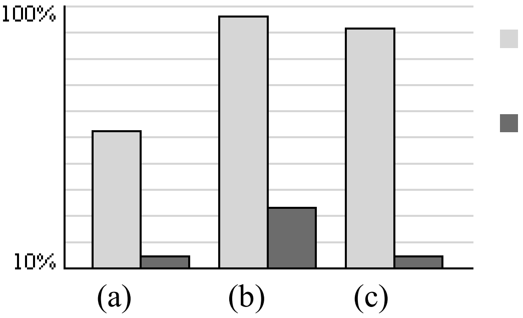

Figure (2) Filter Efficiency versus Different Layers Combinations

We have tested the filter efficiency by inserting a set of spam and legitimate emails to one layer of the filter. Then we re-insert the sample e-mails after combining the previous filters layer with successive layer. In each case filter responses are observed and plotted in figure 2. The gray color represents the spam detected percentage while the black color represents the misclassification percentage. Figure 2-a shows the filter response while testing only the statistical filter layer. This shows a filter efficiency with 62.4% detection percentage and 2% false positive. However, figure 2-b shows an improvement in filter efficiency after combining the Blackhole list layer to the previous layer with 96.6% spam detection percentage. But the false positive has increased to 20%.

Finally, figure 2-c shows a significant improvement in filter efficiency with 93.2% spam detection percentage and 2.3% false positive percentage. These results were obtained after combining additive filter layers consisting of DNS lookup and mail rate control, each having different features to detect spam.

4. CONCLUSION

In this paper, we have presented an integrated solution to protect the mail server from the spam. The main purpose of the proposed filter is to eliminate the spam, or at least reduce the spam rate at the server. Several techniques have been combined to make the filter more efficient in detecting the spam and exhibiting low false positive. Some of these techniques validate the legitimacy of the sender. While the contents of the e-mail are used to classify the mails as spam or legitimate. Using all information in a message (header + body).

A performance measures has been carried out with a set of gathering mails. The results are used to evaluate the performance of the different layers of the filter. A high performance could be observed when combining the proposed filter layers.

REFRENCES

[1] Lazzari Lorenzo, Mari Marco, and Poggi Agostino, "CAFÉ – Collaborative Agents for Filtering E-mails", *Proceedings of the 14th IEEE International Workshops on Enabling Technologies (WETICE'05)*, 2005.

[2] Wang Xiao-Lin, and Cloete Ian, "Learning to Classify E-mail: A Survey", *Proceedings of the Fourth International on Machine Learning and Cybernetics*, pp. 5716-5719, August 2005.

[3] Li Yang, Fang Binxing, Guo Li, and Wang Shen, "Research of a Novel Anti-Spam Technique Based on Usere's Feedbach and Improved Naïve Bayesian Approach",*International Conference on Networking and Services (ICNS'06),* 2006.

[4] Zhag Le, Zhu Jingbo, and Yao Tianshun, "An Evaluation of Statistical Spam Filtering Techniques", *ACM Transactions on Asian Language Information Processing*, Vol. 3, No. 4, Pages 243–269, December 2004.

[5] Androutsopoulos I., Paliouras G., Karkaletsis V., Sakkis G., Spyropoulos C., and Stamatopulos P., " Learning to Filter Spam E-Mail: A Comparison of a Naïve Bayesian and Memory-Based Approach", *Proceedings of the Workshop on Machine Learning and Textual Information Access, 4th Europian Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, 2000.

[6] Saham M., Dumais S., Heckerman D., and Horvitz E., "A Bayesian Approach to Filtering Junk E-Mail," *Proc. AAAI Workshop on Learning for Text Categorization*, Madison, Wisconsin, pp. 55–62, July 1998.

[7] Lai Chih-Chin, and Tsai Ming-Chi, "An Empirical Performance Comparison of Machine Learning Methods for Spam E-mail Categorization", *Proceedings of the Fourth International*

Conference on Hybrid Intelligent Systems (HIS'04), 2004.

[8] Androutsopoulos I., Koutsias J., Chandrinos, K., Paliouras G., and Spyropoulos C., "An Evaluation of Naive Bayesian Anti-Spam Filtering", *Proceedings of the Workshop on Machine Learning in the New Information Age: 11th European on Conference Machine Learning (ECML 2000)*, pages 9–17, 2000.

[9] Islam Md. Rafiqul, Chowdhury Morshed U., and Zhou Wanlei, "An Innovative Spam Filtering Model Based on Support Vector Machine", *Proceedings of the 2005 International Conference on Computational Intelligence for Modeling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)*, 2005.

[10] Xin Wang, Hai-Xin Duan, Tran Q. Anh, and Xue-Nong Li, "Dynamic Rules' Score Adjustment in Spam Filter Using Users' Feedback", *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics*, Guangzhou, 18-21 August 2005.

[11] Clark James, Koprinska Irena, and Poon Josiah, "A Neural Network Based Approach to Automated E-mail Classification", *Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI'03)*, 2003,

[12] Iwanaga M., Tabata T., and Sakurai K., "Evaluation of Anti-Spam Methods Combining Bayesian Filtering and Strong Challenge and Response," *Proc. IASTED International Conference on Communication, Network, and Information Security (CNIS 2003)*, pp. 214–219, 2003.

[13] Brewer Douglas, Thirumalai, Gomadam, and Li Kang, "Towards an Ontology Driven Spam Filter", *Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06)*, 2006.

[14] Damiani Ernesto, Vimercati Sabrina, Paraboschi Stefano, and semarati Pierangela, "P2P-Based Collaborative Spam Detection and Filtewring", *Proceedings of the Fourth International Conference on Peer-to-Peer Computing (P2P'04)*, pp. 176-183, 2004.

[15] Aradhye Hrishikesh B., Myers Gregory K., Herson James A., "Image Analysis for Efficient Categorization of Image-based Spam E-mail", *Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)*, 2005.

[16] Daelemans Walter, Zavrel Jakub, and Der Slot Ko Van, "TiMBL: Tilburg Memory-Based Learner", *Reference Guide, Technical Reporet, Tilburg University*, December 2004. <http://ilk.uvt.nl/downloads/pub/papers>